

The Ontario Veterinary College Hip Certification Program – Assessing inter- and intra-observer repeatability and comparison of findings to those of the Orthopedic Foundation for Animals

Heather J. Chalmers, Stephanie Nykamp, Assaf Lerer

Abstract – In Canada, the Ontario Veterinary College (OVC) has offered radiographic screening for hip dysplasia for many years, but there are other options for this service including the Orthopedic Foundation for Animals (OFA). There are some differences between the OFA and the OVC methods, and this study compares the OVC and OFA hip certification results in 37 dogs. There was good agreement between the two programs but in some instances there was a difference in the pass/fail status of a dog. Neither the OFA nor the OVC was more likely to fail or pass a given dog. The repeatability of the OVC results was assessed by both inter- and intra-observer comparisons in 100 dogs. There was at least 86% agreement among and within radiologists, but in 5 cases the disagreement resulted in a difference in the pass/fail status of the dog. These results illustrate the inherent variation in radiographic hip evaluation and highlight the importance of consensus grading practices to improve the accuracy of hip evaluation.

Résumé – Programme de certification des hanches de l'Ontario Veterinary College — Évaluation de la reproductibilité inter- et intra-observateur et comparaison des résultats à ceux de l'Orthopedic Foundation for Animals. Au Canada, l'Ontario Veterinary College (OVC) offre le dépistage radiographique de la dysplasie de la hanche depuis de nombreuses années, mais il y a d'autres options pour ce service, incluant l'Orthopedic Foundation for Animals (OFA). Il y a certaines différences entre les méthodes de l'OFA et de l'OVC et cette étude compare les résultats de certification de la hanche de l'OVC et de l'OFA chez 37 chiens. Il y avait une bonne concordance entre les deux programmes, mais dans certains cas, il y avait une différence au niveau du statut d'échec-réussite d'un chien. Ni l'OFA ni l'OVC ne présentait une probabilité accrue de donner un résultat d'échec ou de réussite à un chien particulier. La reproductibilité des résultats de l'OVC a été évaluée par des comparaisons inter- et intra-observateur chez 100 chiens. Il y avait au moins 86 % de concordance entre et parmi les radiologistes, mais dans 5 cas, la discordance s'est traduite par une différence du statut de réussite et d'échec chez le chien. Les résultats illustrent la variation inhérente à l'évaluation radiographique de la hanche et souligne l'importance de pratiques de classification par consensus afin d'améliorer l'exactitude de l'évaluation de la hanche.

(Traduit par Isabelle Vallières)

Can Vet J 2013;54:42–46

Introduction

Canine hip dysplasia (CHD) is an important and common problem for pet and breeding dogs in Canada and worldwide. The disease has a known genetic basis and is heritable; however, environmental factors impact the phenotypic expression and the severity of the disorder in affected individuals (1–6). The disease is progressive, and once initiated can result

in reduced range of motion, pain, and lameness (7,8). Early diagnosis of hip dysplasia is essential to facilitate early management strategies and to prevent breeding of affected individuals.

Changes in hip joint congruity and stability may be detected on orthopedic examination including the Ortolani sign and Barden test (9–11); however, radiographic evidence is necessary to determine the nature and severity of the hip dysplasia, to assess for secondary degenerative joint disease, and facilitate treatment planning in affected dogs. Dogs with hip dysplasia may exhibit a range of radiographic findings, with abnormalities including variable degrees of hip joint incongruity, subluxation, and degenerative joint disease (12).

Screening programs for hip dysplasia have traditionally involved radiographic assessment of the hip joints of young adult dogs, and there are many such programs available worldwide. These include the Orthopedic Foundation for Animals (OFA) program, Pennsylvania Hip Improvement Program

Department of Clinical Studies, Ontario Veterinary College, University of Guelph, Guelph, Ontario N1G 2W1.

Address all correspondence to Dr. Heather J. Chalmers; e-mail: heather.chalmers@uoguelph.ca

Use of this article is limited to a single copy for personal study. Anyone interested in obtaining reprints should contact the CVMA office (hbroughton@cvma-acmv.org) for additional copies or permission to use this material elsewhere.

(PennHIP), Fédération Cynologique Internationale (FCI), the British Veterinary Association hip scheme, the Ontario Veterinary College Hip Certification program, and numerous other breed specific and/or country specific programs. While these organizations share the common goals of facilitating early diagnosis and reducing the incidence of CHD in the dog population, each organization has its own specific methods and grading schemes. As a result, it can be difficult for breeders, pet owners, and even veterinarians to compare results between the various programs. It is our impression that the most commonly used programs in Canada are the OVC Hip and Elbow Certification Program (OVC-HCP), the PennHIP, and the OFA program. Of these, the OVC-HCP and the OFA have similar systems requiring a ventrodorsal hip extended radiograph of the pelvis (13). In addition to the traditional view, PennHIP utilizes a dynamic radiograph series in which the passive hip laxity is calculated. In order to perform the PennHIP test, an individual must acquire special training leading to certification (14).

The OFA is a non-profit organization founded in 1966. The OFA maintains the world's largest database of radiographic hip evaluations. Radiographs submitted to the OFA are assessed by 3 board-certified radiologists; once submitted, the radiographs are not returned. The OFA hip grading method consists of 3 main outcomes; pass, borderline, fail. Dogs which receive a passing score are assigned to 1 of 3 categories: Excellent, Good, or Fair. Dogs which receive a fail score are assigned 1 of 3 grades: Grade 1, 2, or 3, which represent increasing severity. The OFA provides certification services on dogs that are a minimum of 24 mo of age, and also offers preliminary assessment of younger dogs.

The OVC-HCP has been active for over 25 y, and certifies between 1200 and 5000 dogs per year. The program was started largely in response to the need for a Canadian option for breeders. At that time, breeders found that shipping radiographs to the United States for OFA evaluation resulted in long delays and increased costs associated with shipping and exchange rates. In order to address this need, the OVC began to offer hip and elbow certification using an OFA style ventrodorsal hip extended radiograph of the pelvis, which was evaluated by a board-certified radiologist. The program has other distinctions when compared to the OFA including that only 1 radiologist interprets each radiograph, dogs may be certified at a minimum of 18 months of age (with preliminary results available in younger dogs), and the original radiographs are returned to the client. From these beginnings, the program has evolved to include an online searchable database, and at present radiographs submitted to the OVC-HCP are assessed by 1 of 2 board-certified radiologists. The OVC hip grading system has 2 main outcomes: Pass and Fail. Dogs that receive a fail are assigned to Grade 1, 2, 3, or 4, representing increasing severity of CHD (Table 1). Despite being the only Canadian certification program and being a historical favorite of many Canadian breeders, the OVC method of assessing hips has not been scientifically evaluated to date.

Within the North American purebred dog industry, there is considerable breeding of dogs between provinces and states, and even internationally. Establishing the orthopedic health of a dog is an essential part of responsible breeding practice. Often,

Table 1. The Ontario Veterinary College Hip Certification Program grading system. All radiographs are ventrodorsal (hip extended) view and are evaluated by a board-certified veterinary radiologist. The radiographs are systematically assessed for all aspects of hip joint conformation including parallelism between the femoral head and acetabulum, shape of the femoral head and acetabulum, coverage of the femoral head by the acetabulum, evidence of subluxation, and evidence of degenerative joint disease. The results of the systematic evaluation for each dog are summarized by assigning a semi-quantitative grade.

OVC assessment	Typical findings	Result/Outcome
PASS		
Normal	No radiographic evidence of hip dysplasia, including normal hip conformation and no evidence of degenerative changes.	Certificate issued.
FAIL		
Grade I	Minimal or mild change is present including incongruity of hip joint, or reduced coverage of femoral head, or enthesiophyte on femoral neck.	No certificate, a report outlining radiographic findings and stating grade is issued.
Grade II	Moderate changes or more than one mild change are present.	
Grade III	More than one moderate change is present.	
Grade IV	Severe changes are present.	

breeders and veterinarians use 1 program and become more familiar with interpreting results from that program than from others. This can create difficulty when a dog from 1 owner for which the OVC-HCP certificate is available plans a mating with a dog from another region for which the OFA results are available. It is therefore desirable to compare results between programs to assist breeders and veterinarians with decision-making. In some cases, it has been necessary to re-submit radiographs of a dog to a second program in order to satisfy all parties that the intended mating involves only orthopedically sound individuals. This is associated with increased cost for breeders because of submission fees and in some cases additional radiographs. Because the OFA and OVC-HCP utilize the same radiographic view and similar assessment criteria, it is desirable to establish some basic guidelines that can be used for extrapolation of results between the OFA scoring and the OVC scoring systems.

The goals of the current study were to i) establish the repeatability of the OVC-HCP method by establishing the inter- and intra-observer repeatability of this method, and ii) determine the agreement between the OFA and the OVC-HCP results by comparing the findings between the 2 programs in a small subset of dogs.

Materials and methods

Inter- and intra-examiner repeatability

A sample of 100 sequential cases submitted to the OVC hip screening program were enrolled in the study. Cases were excluded from the study if the radiographs were deemed to be of non-diagnostic quality or if the dog was less than 18 mo of age; these cases were replaced with the next case submission until 100 cases were enrolled. These 100 cases were randomized and

each radiologist completed the evaluation for each dog 3 times with a minimum of 24 h between readings (total of 6 evaluations per dog). For each reading, the radiologist was blinded to the previous results and the results of the other radiologist. In addition to performing the hip certification procedure, the radiologist was also asked if they would seek a second opinion on the case.

Method comparison study

The comparison between the OVC method and the OFA method was performed on 37 subjects, which were different from the 100 dogs previously described because the method comparison portion of the study was done after the repeatability portion. These subjects were taken as a subsample of dogs submitted to the OVC-HCP, with the inclusion criterion for participation in the study being dogs of at least 2 y of age and client consent during the study period (December 2010 — July 2011). Participation was solicited by posting the research project information on the OVC-HCP Web site and also by direct telephone or e-mail communication to clinics with a high submission rate to the program. Each radiograph was routinely read and graded by the radiologist on duty without knowledge of enrollment in the study. The assessment method is familiar to both radiologists participating in the study and is unmodified from the assessment method for routine submissions. Based on the radiograph, dogs were assigned to 1 of the OVC-HCP grade categories (Table 1).

The same radiographs were scanned and digitally submitted to the OFA for routine evaluation. This was performed with consent of the OFA administration, and the OFA radiologist reading the cases was unaware of the study. The OFA results were collected and categorized according to the standard OFA scale (http://www.offa.org/hd_grades.html). The OFA evaluation was unmodified from a routine submission except that only 1 radiologist read each case.

Statistical methods

For the quantification of inter-observer repeatability within the OVC method, the intra-class correlation coefficient was calculated. Standard 2×2 tables were constructed for pass/fail between the 2 observers and for the need for a second opinion, and basic summary statistics were calculated. For the intra-observer repeatability, the kappa statistic was calculated with the most common response of each radiologist being used for the comparison. With regards to the choice to seek a second opinion, the kappa statistic was also calculated to compare the agreement between the 2 radiologists about the need for a second opinion. For the ordinal data, the OFA and OVC-HCP results were compared using a kappa statistic. For both the OVC-HCP and the OFA, the results were dichotomized for part of the analysis to determine the differences between the hip status of “pass” or “fail” between the 2 programs. For the OFA, the grades were dichotomized such that OFA grade excellent, good, fair were considered equivalent to “pass” and OFA borderline, mild, moderate, and severe were considered equivalent to “fail.” For the OVC-HCP, the dogs that passed and received a certification were considered to “pass” and the dogs that did not

pass and received a grade of I-IV hip dysplasia were considered to “fail.” The McNemar’s test was used to assess for overall bias of the 2 programs, or tendency of 1 program to rate consistently higher or lower than the other. For both study populations, the mean age, and sex distribution were compared to those for the general OVC-HCP submissions obtained via database query.

Results

Repeatability study

The 100 dogs in this portion of the study consisted of 68 females (64 intact and 4 spayed) and 32 males (29 intact and 3 castrated) with a mean age of 21.2 mo (range: 18 to 109 mo). Breeds represented included Labrador retrievers ($n = 29$), German shepherd dogs ($n = 20$), golden retrievers ($n = 14$), Bernese mountain dogs ($n = 7$), rottweillers ($n = 3$), poodle-standard berger allemande ($n = 2$), ($n = 2$), Labrador retriever crosses ($n = 8$), and 1 dog of each of 15 other breeds (löwchen, American cocker spaniel, bull mastiff, Alaskan malamute, Samoyed, Portugese water dog, Irish setter, Newfoundland, border terrier, rough collie, smooth collie, weimaraner, German wire-haired pointer, Australian kelpie, great dane, greater Swiss mountain dog). The mean age and gender distribution did not differ from the means of the general submissions for the year of the study.

For the 100 dogs evaluated, the grade assigned to the dog was the same for all 6 observations a total of 86 times. Of the 14% of the time that there was any disagreement about the grade of a dog, this resulted in disagreement about the pass/fail status of the dog in 5/14 cases (35.7%). In 12/14 (85.7%) of the cases where there was disagreement regarding the grade, 1 or both radiologists stated that they would seek a second opinion. The most common type of disagreement either between or within observers was between grade I and grade II status (64.2% of all disagreements), and of the 5 cases in which the disagreement resulted in a difference in pass/fail status 4/5 were differences between grade I and normal and 1/5 was a difference between grade II and normal. The kappa for seeking a second opinion, representing the instances in which both radiologists sought a second opinion on the same case, was 0.26. One radiologist was significantly more likely to seek a second opinion than the other ($P < 0.05$). The combined occurrence of seeking a second opinion was 19/100 (19%), with 14/19 (73.6%) being sought by 1 radiologist and 5 being sought by the other radiologist. There was no difference between the 2 radiologists in the frequency of assigning the various grades. The kappa for inter-observer agreement for each observation was observation 1: 0.742, observation 2: 0.770, and observation 3: 0.787 (Fleiss-Cohen weighted kappa, squared weighting $P < 0.0001$ for all 3 observations). For the intra-observer agreement, the intra-class correlation coefficient was 0.78.

Method comparison study

The 37 dogs consisted of 12 intact males and 25 intact females with a mean age of 36.7 mo (range from 24 to 74 mo). The mean age was significantly different from the mean age of the general submissions of the year of the study ($P = 0.002$). Breeds represented included golden retrievers ($n = 4$), Labrador

Table 2. Results of OVC-HCP and OFA assessment in 37 dogs submitted for certification

	Results of OFA assessment					
	Excellent	Good	Fair	Borderline	Mild	Moderate
Results of OVC-HCP assessment						
Pass	2	25	2	0	0	0
Grade I	0	2	2	0	0	0
Grade II	0	0	0	0	2	0
Grade III	0	0	0	0	1	2

retrievers ($n = 3$), Shetland sheepdogs ($n = 3$), Irish water spaniels ($n = 2$), Alaskan malamutes ($n = 2$), Nova Scotia duck tolling retrievers ($n = 2$), and 1 of each of 21 other medium and large breeds.

The OVC radiologists gave 29/37 dogs a pass and 8/37 dogs a fail. Of the dogs that received a fail score from the OVC-HCP, 4 dogs received a Grade 1 score, 2 dogs received a Grade 2, and 2 dogs received a Grade 3 score (Table 2). No dog received Grade 4 score. The OFA gave 33/37 dogs a passing score and 4/37 dogs a failing score. No dog was given a borderline score. Out of the dogs that received a passing score, 2 dogs received an excellent score, 27 dogs received a good score, and 4 dogs received a fair score. From the dogs that received an OVC-HCP fail score, 3 dogs received Grade 1, and 1 dog received Grade 2. All the dogs which received a fail score by the OFA also received a fail score by the OVC, 4 dogs that received a passing score by the OFA received a fail score by the OVC; 2/4 of these dogs received an OFA grade score of good, and 2/4 dogs received an OFA grade score of fair.

When the dichotomized pass/fail status of the dog is considered, the kappa between OVC-HCP and OFA = 0.6105 [$P = 0.0011$, 95% confidence interval (CI): 0.2785 to 0.9425]. McNemar's test of bias was not significant ($P = 0.5$). The odds ratio for passing OFA and failing OVC-HCP is 5.3 (median unbiased estimate, 95% CI: 0.66 to 157.5).

Discussion

Early and accurate screening for CHD is an essential component of canine breeding programs and several screening options are available to Canadian veterinarians and breeders. Radiographic assessment and measurements form the foundation of virtually all of these programs and the results are used by veterinarians and breeders to advance the orthopedic status of dogs through responsible breeding. In this study, we compared the hip joint grading score of the OVC-HCP and the OFA program. Both methods are based on a semi-quantitative evaluation of an extended coxofemoral joint in ventrodorsal view and are commonly used in Canada.

Overall, the correlation between the grading score provided by the 2 radiologists at OVC is categorized as good, as reflected by the ICC of 0.78 (15). There have been multiple studies assessing the repeatability of the various hip dysplasia scoring methods. For the PennHIP method, the ICC was reported by the PennHIP researchers to be between 0.85 and 0.94 (16) and more recently was reported to be as high as 0.96 (17). These authors have concluded that the PennHIP scores obtained by various trained observers could be considered interchangeable

(17). A study that focused on agreement of the FCI method showed that inexperienced observers have poor agreement (ICC up to 0.44), while more experienced observers have good agreement (ICC up to 0.72) (18). In human medicine, the ICC for various radiographic hip scores range from poor (0.49) to excellent (0.97), with most studies reporting good repeatability (19,20). The wide range of results for agreement and the variable interpretation of what should be considered acceptable agreement complicate the interpretation of the current results.

One important difference between the OFA and the OVC-HCP is that with the OFA program, each radiograph is evaluated by 3 radiologists while at the OVC only 1 of 2 radiologists performs the evaluation on any given dog although a second opinion may be sought informally. As such, it is important to establish the inter-observer repeatability of the OVC-HCP scoring to ensure that the results are reliable between the 2 radiologists. Of the 100 dogs evaluated here, the radiologists gave the same score 86% of the time. Of the 14 dogs in which the 2 radiologists did not agree about hip score, this resulted in a difference in the pass/fail status of 4 dogs. However, in all of these cases 1 or both radiologists indicated that she/he would seek a second opinion before deciding on the final score for the dog. This reinforces the importance of the practice of seeking a second opinion and implies that there may be some benefit in implementing a more formal requirement for a second opinion as is routine with the OFA. This may include measures such as requiring 2 observers for each submission, or formalizing the process of seeking a second opinion.

Factors that affect inter- and intra-observer repeatability of OVC-HCP hip certification results could include technical quality of the radiographs (20), the individual set points/criteria of a given observer, experience of the observer, and bias relating to knowledge of information such as breed or age. In the present study, only radiographs of adequate technical quality are accepted by OVC-HCP and were included in the study. Additionally, the 2 observers are of similar experience and training and were blinded to the signalment information for the dogs. The observed differences are therefore most likely attributable to differences in the opinion and set points of the 2 radiologists. This type of difference of opinion is well-recognized in medical imaging as a common source of disagreement.

The comparison of various hip screening methods has been widely studied by many parties, with the aim of establishing which test is a more accurate predictor of the long-term hip status of the dog in order to make a recommendation of 1 program over another. In the current study, this was not the goal and the 2 programs are not being compared in order to determine

which is superior, but rather in order to allow some understanding of differences between the programs. A gold standard in establishing the true hip status of a given dog was not available in this study; therefore, the results from the 2 programs cannot be compared to each other for assessing which results are more correct relative to a gold standard. Rather, these comparisons are made to establish the correlations between the 2 grading schemes in order to broaden current understanding of both programs and facilitate extrapolation of results between the 2 programs where necessary.

Guidelines for extrapolation between the 2 programs, recognizing the inherent limitations, would be desirable. It is important to note that the current study assessed only dogs 24 mo of age or older. The OVC-HCP performs certifications on dogs as young as 18 mo, and the OFA performs preliminary certification on dogs younger than 24 mo, but the comparison between results for dogs younger than 24 mo has not been evaluated here. The results of this study, therefore, cannot be applied to younger dogs. While the current study identified a minority of dogs that received a “pass” grade at the OFA and a “fail” grade from OVC-HCP, the McNemar’s test of bias was not significant indicating that there is no preferential direction to the disagreement between the OFA and the OVC-HCP. Neither program is more likely to pass or fail a given dog, and there is no bias or differential rate of pass or fail between the 2 programs that could be detected in this sample. However, based on the odds ratio, the odds of the OFA passing a dog that was failed by OVC-HCP is 5.3× higher than the odds of a dog failing the OFA and passing the OVC-HCP assessment. Another important distinction between the 2 programs is that the OFA provides a grade within both passing and failing cases. This may be very useful for breeders, who could decide what level within a passing grade is acceptable for their own standards. For example, a dog receiving an OFA score of “fair” is considered to be a pass in the current project; however, many breeders would make the commitment that only dogs of good or excellent status are used for breeding. The provision of a grade of normalcy within the OFA system is likely an advantage over the OVC system, in which the radiologist makes a pass/fail distinction and a grade is only provided for the failures.

Among responsible breeders and veterinarians there is continued emphasis on the prevention of heritable diseases such as CHD through good breeding practices. The hip certification programs play a key role in providing accurate results about hip dysplasia and other heritable conditions with the common goal of improving the genetic health of dogs. The current study provides some validation of the repeatability of the OVC-HCP and offers some examples of how the grading schemes between the OVC-HCP and the OFA can differ, and demonstrates that results from 1 program consistently compare to the other without bias. The findings support that there is consistency within

and between radiologists participating in the program and that where there is lack of agreement, this rarely results in a change in the pass/fail status of a dog. The needs of veterinarians and the breeding industry are best served by those programs which provide interpretation in an accurate and repeatable manner and the current work demonstrates that both programs are suitable for this purpose.

CVJ

References

1. Chase K, Lawler DF, Carrier DR, Lark KG. Genetic regulation of osteoarthritis: A QTL regulating cranial and caudal acetabular osteophyte formation in the hip joint of the dog (*Canis familiaris*). *Am J Med Genet A* 2005;135:334–335.
2. Chase K, Lawler DF, Adler FR, et al. Bilaterally asymmetric effects of quantitative trait loci (QTLs) that affect laxity in the right versus left hip joints of dogs. *Am J Med Genet* 2006;124:239–247.
3. Lust G, Rendano VT, Summer BA. Canine hip dysplasia: Concepts and diagnosis. *J Am Vet Med Assoc* 1985;187:638–640.
4. Hedhammar A, Wu FM, Krook L, et al. Overnutrition and skeletal disease. *1974;64Suppl 5:32–45*.
5. Corley EA. Role of the orthopedic foundation for animals in the control of canine hip dysplasia. *Vet Clin North Am Small Anim Pract* 1992;22:579–93.
6. Kaneene JB, Mostosky UV, Miller R. Update of a retrospective cohort study of changes in hip joint phenotype of dogs evaluated by the OFA in the United States, 1989–2003. *Vet Surg* 2009;38:398–405.
7. Alexander JW. The pathogenesis of canine hip dysplasia. *Vet Clin North Am Small Anim Pract* 1992;22:503–511.
8. Lust G. An overview of the pathogenesis of canine hip dysplasia. *J Am Vet Med Assoc* 1997;210:1443–1445.
9. Lust G, Summers BA. Early, asymptomatic stage of degenerative joint disease in canine hip joints. *1981;42:1849–1855*.
10. Lust G, Beilman WT, Dueland R, et al. Intra-articular volume and hip joint instability in dogs with hip dysplasia. *J Bone Joint Surg Am* 1980;62:576–582.
11. Lust G, Beilman WT, Rendano VT. A relationship between degree of laxity and synovial fluid volume in coxofemoral joints of dogs predisposed for hip dysplasia. *Am J Vet Res* 1980;41:55–60.
12. Slatter DH, ed. *Textbook of Small Animal Surgery*. 3rd ed. Philadelphia, Pennsylvania: Elsevier Science, 2002:2009–2019.
13. Rendano VT, Ryan G. Canine hip dysplasia evaluation. *Vet Radiol* 1985;26:170–186.
14. Fluckiger MA, Friedrick GA, Binder HA. Radiographic stress technique for evaluation of coxofemoral joint laxity in dogs. *Vet Surg* 1999;28:1–9.
15. Portney LG, Watkins MP. *Foundations of clinical research applications to practice*. Upper Saddle River, New Jersey: Prentice Hall, 2000: 560–567.
16. Smith GK, LaFond E, Gregor TP, Lawler DF, Nie RC. Within- and between-examiner repeatability of distraction indices of the hip joints in dogs. *Am J Vet Res* 1997;58:1076–1077.
17. Ginja MM, Ferreira AJ, Silvestre M, Gonzalo-Orden JM, Llorens-Pena MP. Repeatability and reproducibility of distraction indices in PennHIP examinations of the hip joint in dogs. *Acta Vet Hung* 2006;54:387–392.
18. Verhoeven GEC, Coopman F, Duchateau L, et al. Interobserver agreement on the assessability of standard ventrodorsal hip-extended radiographs and its effects on agreement in the diagnosis of canine hip dysplasia and on routine FCI scoring. *Vet Radiol Ultrasound* 2009; 50:259–263.
19. Mast NH, Impellizzeri F, Keller S, Leunig M. Reliability and agreement of measures used in radiographic evaluation of the adult hip. *Clin Orthop Relat Res* 2011;469:188–199.
20. Clohisy JC, Carlisle JC, Trousdale R, et al. Radiographic evaluation of the hip has limited reliability. *Clin Orthop Relat Res* 2009;467:666–675.